

Review

Peer review and the process of publishing of adverse drug event reports

Steven B. Karch MD *

*P.O. Box 5139, Berkeley, CA 94705, United States*Received 20 December 2005; received in revised form 18 January 2006; accepted 10 February 2006
Available online 22 August 2006

Abstract

Most doctors finish their careers without ever having published a scientific paper. The small percent that do mostly end up writing “case reports”. But, unless the report describes some truly extraordinary event, very little ever comes of it, and their publication may do more harm than good. The problem is particularly acute in cases of alleged drug toxicity. Case reports are incomplete, uncontrolled, retrospective, lack operational criteria for identifying when an adverse event has actually occurred, and resemble nothing so much as hearsay evidence, a type of evidence that is prohibited in all courts in all of industrialized societies. When a journal, even a highly respected one, decides to publish hearsay, readers are utterly reliant on the integrity of the journal, the author, and the peer review process; such reliance may not always be warranted. Some recent examples of process failure are provided, as well as some suggestions about possible remedies, including the use of pharmaco-vigilance algorithms.

© 2006 Elsevier Ltd and AFP. All rights reserved.

Keywords: Case report; Case series; Hearsay evidence; Toxic torts; Naranjo; Randomized clinical trials; Junk science

1. Introduction

In 1993, the US Supreme Court ruled that judges must act as gatekeepers, excluding from court any sort of unreliable evidence. Similar rulings do not as yet apply in either the United Kingdom or the EU. Still, many would insist that judges in those countries share a similar obligation. Those of us who practice in the courts tend to forget that judges are not the only ones who should be acting as gatekeepers. Journal editors and reviewers have a similar duty to halt the dissemination of junk science.

The curriculum of most medical schools allows little time for instruction in the principles of scientific method, and even less for instruction in the fundamentals of legal causation. The medical curriculum focuses, quite rightly, on curing patients, not going to court, but this focus has a cost. The average medical doctor is not a trained researcher and is not likely to know “the rules of the game”. Neither is the judge. This may explain why case-reports are almost always admitted as evidence in court.

Questions such as how many babies have really died from being shaken, or how many positive finger print identifications were truly positive, or how many cocaine addicts in police custody were really killed by “positional asphyxia”, or how many dieters have actually died from ephedrine toxicity have no clear answers. Most of what we know and say about these subjects is derived from anecdotal case reports, not controlled scientific studies. Nearly a quarter century ago, transplant surgeon Michael DeBakey wrote that Adverse Event Reports (AERs) were “widely abused”, and their quality “mediocre”.¹ There has been little progress in the last quarter century.

In 2003 Kelly analyzed a total of 1520 case reports, published over a 20-year period (from 1978 to 1997).² The sample was limited to AERs describing only highly significant events (death, permanent disability, or life threatening reactions). Each case report was analyzed for the presence of 15 different variables, including age, gender, presence of other diseases, drug abuse history, liver and renal function. Age, sex, and outcome were mentioned in 90% of the reports, but information about the other 12 variables was reported in less than 25%. Renal function was mentioned in only 6% of the cases. Liver function, a major issue in

* Tel.: +1 510 849 0623; fax: +1 510 849 0958.

E-mail address: Skarch@sonic.net.

1. Strength of relationship
2. Consistency of relationship
3. Specificity (is the effect also associated with other agents?)
4. Temporality (did the effect occur after exposure)
5. Biological Gradient (Dose/Response Curve)
6. Plausibility (Do not dismiss an observation at odds with what is known)
7. Coherence (is the finding at odds with current knowledge?)
8. Experiment (is their experimental data confirming the finding)
9. Analogy (are the effects similar to that of another drug)

Fig. 1. Bradford Hill criteria.

the assessment of drug toxicity, was described in less than 1% of the reports. Blood level measurements were reported in only 14% of cases, and formal methods for casualty assessment were applied in less than 1%. Nonetheless, possible mechanisms of toxicity (authors sometimes referred to the process as an assessment of “biologic plausibility”, ignoring all the rest of Bradford Hill’s criteria²⁰ – see Fig. 1) were discussed by more than 90% of the authors. Kelly also found that none of the 12 most frequent publishers of AERs (the list includes *Archives of Internal Medicine* and the *New England Journal of Medicine*) only one, the *Annals of Pharmacotherapy*, actually required authors to submit an objective casualty assessment.

In spite of these well-documented flaws, case reports are still admitted as evidence in courts, and it is time the forensic community did something to remedy the situation. A good way to start might be by more closely vetting, or even limiting the number of AERs published, the kind of evidence often relied upon in “toxic tort” litigation. This sort of litigation is increasingly common, and so is reliance upon AERs. The only good thing to be said about the situation is there is no shortage of real, illustrative cases with which to illustrate the problem.

2. Case reports of AERs

Most doctors finish their careers without ever having published a scientific paper; the small percentage that do mostly end up writing case reports. Doctors in training, especially those with academic aspirations, are encouraged to “write a paper”. Not much effort is required to produce a case report; compile a brief account of the patient’s history, described the odd feature observed, review similar cases that have been reported in the past, and offer some speculative theory about why this particular drug caused this particular reaction. Unless the report describes some truly extraordinary event it, very little ever comes of it, except, of course, an addition to the individual’s curriculum vitae.

If clusters of reports, describing similar drug reactions are published, the cluster would legitimately provide grounds for hypothesis-testing experiments. One report of eosinophilic myocarditis in a cocaine user would elicit little interest in the medical community. Does cocaine cause eosinophilic myocarditis?³ Perhaps, maybe, under some circumstances. But there have only been three reports of eosinophilic myocarditis in cocaine users published in the last 15 years, so there is little interest in the subject, no

funding, and no hypothesis-testing experiments. There may or may not be a relationship cocaine usage and the occurrence of eosinophilic myocarditis, but one thing is for sure: the entity will be described in all forthcoming review papers about cocaine and heart disease.

The same concerns apply even when a group of case reports are combined and published as a “case series”. There is no strength in numbers. A case series suffers from all of the same flaws as a single case report or AER. These reports are incomplete, uncontrolled, retrospective, and lack operational criteria for identifying when an adverse event has actually occurred. No matter how many cases are included in the series, each case, individually, is subject to the same weaknesses and flaws as all of the others.^{2,4}

Case reports are subject to the potential bias of authors which may be unintended. The two case reports summarized below were published late in 2005.^{5,6} The authors clearly believe their reports demonstrate a cause and effect relationship, and no data that might conflict with their conclusion appeared to be referred to.

A 45-year-old woman with a history of migraine, who complained of a headache, took two aspirins, and sustained a cardiac arrest. She was successful resuscitated, but died after six days of ventilator and vasopressor (i.e., infusion of two different catecholamines, dopamine and norepinephrine) support. The paper does not specify the dosage of vasopressors administered prior to death. Medical history disclosed the decedent had been taking an unspecified dose of an ephedrine-containing supplement, for an unspecified period of time. The decedent also had a 25-pack-year smoking history, and was using a nicotine inhaler at the time of her arrest. She was also taking an unspecified dose of Prozac for an unspecified time (the Prozac package insert, and the label on the herbal supplement both contains a warning about possible dangerous interaction between ephedrine and reuptake inhibitors). Toxicology screening at the time of admission was positive for cannabinoids but was said to be negative for all other drugs, though the manner of drug testing was not specified and confirmatory testing not performed. Arteriography was not performed. Autopsy disclosed non-specific degenerative changes in the heart, including evidence of apoptosis, contraction band necrosis and ischemia.

The authors do not explain that all of the myocardial changes could be the result of continuous administration of vasopressors during a week of life support.^{7,8} DNA testing for viruses that might cause myocarditis was not performed, even though it is clear that cardiotropic viruses may infect the heart, cause acute coronary syndromes, and leave no histological evidence.⁹ Blood ephedrine levels were never measured, and so it is not even clear whether the decedent took ephedrine or, if she did, how much. The dose-response relationship, which is a central concept in both clinical and forensic medicine, was never considered.¹⁰ Hair testing that would have disclosed evidence of recent

abuse of other drugs abuse (or confirmed ephedrine use) was not performed. The fact that the decedent was a poly-drug user was never discussed. Alternate causation was never even considered.

Yet, in spite of all these factors, the authors conclude, “*Healthcare professionals are therefore urged to warn their patients about the risk of serious adverse effects, which may follow ephedra intake*”. The paper contained no warnings against smoking, or taking products that carry warnings about potentially dangerous interactions. Some might interpret the disconnect between evidence proffered and conclusion reached as agenda driven.

Reports of similar concern regularly appear in other mainstream journals. In December of last year the *Lancet* published a paper entitled *Case Report: Consequences of ephedra use in an athlete*. It described the case of a 37-year-old self-admitted intravenous steroid abuser who presented with atrial fibrillation, apparent cardiomyopathy, and an embolic stroke.¹¹ There was no Holter monitoring, no cerebral angiogram, no myocardial biopsy, and no hair testing. But screening toxicology testing was done, and it was negative for amphetamines (ephedra cross reacts in these tests, strongly suggesting that it was not present). The authors of the report tell us that “*Ephedra seems to predispose to both ischaemic and haemorrhagic stroke*”, but provide no references except to other case reports. They also fail to mention the more than 50 RCTs, involving thousands of patients, where ephedrine did not cause stroke or embolism. Nor did they mention the findings of a recent large epidemiologic study that concluded there was no relationship.¹²

3. Case series

In the United States women have been charged with homicide for having transmitted methamphetamine in their mothers milk, or across the placenta.^{13,14} More than half a dozen of these cases have been tried to date, many in California and, most recently, in Hawaii. Since no randomized clinical trial has ever demonstrated fatal or, for that matter, any sort of injury, after human fetal or infant methamphetamine exposure, prosecution experts almost always cite a paper entitled, “*Fetal and infant deaths associated with maternal methamphetamine abuse*”. The paper is actually a “case series”, reporting blood and tissue levels in six still born infants.¹⁵ If levels in the alleged murder victim fall within the ranges observed in the six deaths reported in the paper, the prosecution’s expert will argue that drugs caused the death. The line of reasoning may seem plausible to the jury since it establishes a “range”, and number ranges are easier for a jury to understand than histological changes in the cardiac conduction system, or viral genome sequencing. The problem here is that none of the six reported infants died of methamphetamine poisoning; their deaths had all been attributed to other causes and the presence of methamphetamine was considered an incidental finding by the pathologists who signed the death certificates.

4. Why case reports are unreliable indicators

As demonstrated by the recent Merck-New England Journal debacle, the peer-review procedure is vulnerable in a number of areas, and these vulnerabilities are not limited to case reports describing adverse drug reactions. The problem of “background noise” is, perhaps, the most vexatious and difficult to solve.

4.1. Background “Noise”

Some case reports are more convincing than others. A report describing the occurrence of a very rare event, such as Stevens-Johnson syndrome, in the user of a new medication, must be taken very seriously. A report describing the occurrence of a common disorder such as myocardial infarction or cardiac arrest, in a subject taking a common medication, such as acetaminophen, is much less cause for concern. Survey data suggest that 20% of all Americans, or nearly 60 million people used acetaminophen in 2004.¹⁶ The accepted rate for acute myocardial infarction in the US is 0.52% per year. Even assuming that only one half of the users were seriously at risk for infarction (age >35 years), more than 312,000 (0.52% of 30 million episodes of myocardial infarction) would have occurred in individuals taking acetaminophen each year. The accepted rate for cardiac arrest resulting in sudden death in the US, is 0.1% per year, which means that another 30,000 acetaminophen users would have died suddenly, if only by chance. The rates for stroke are much higher than sudden cardiac death, with nearly 0.25% per year, or 75,000 episodes, bringing the total number of cardiovascular events occurring in American acetaminophen users to well over half a million per year.¹⁷

The numbers explain why no competent reviewer would be likely to accept a case report describing an episode of myocardial infarction in an acetaminophen user. The problem for readers and editors alike is that similar computations are rarely performed, often because the degree of exposure (drug sales, percent contamination, at risk behaviors, etc.), best described as the numerator, is rarely known with certainty. Alternatively, it could be that the authors never state the numerator because it might spoil the argument they are trying to advance or, the reviewer may be unaware of the magnitude of the background “noise”.

4.2. Data integrity

Until very recently it was unusual for journal reviewers to see, let alone examine in detail, the underlying data. The pattern is changing, largely because of recent scandals at some prominent journals. Journals, such as *Nature*, now routinely provide raw data from clinical trials and laboratory studies online for examination by readers (and also to save space in the journal). However, reviewers rarely, if ever, reexamine the underlying data that forms the basis for case reports and case series. And, as the executive editor

of the *New England Journal of Medicine* (NEJM) recently discovered while being deposited in the Vioxx litigation,¹⁸ not all researchers submit all their data for inspection. According to an editorial in the December 29th issue, NEJM's editors have since discovered that other relevant data had been deleted from the paper, forcing them to conclude, “Taken together, these inaccuracies and deletions call into question the integrity of the data on adverse cardiovascular events in this article. We have asked the authors to submit a correction to the Journal”.¹⁹

The NEJM problems were unearthed by expert epidemiologists and statisticians employed by that journal. But what if the person reading the article is untrained in evidence-based medicine, and has never heard of Bradford Hill or his criteria?²⁰ In that situation, all the reader can do is assume the reviewers have done their job. And if it is a case report, and not a randomised controlled trial at issue, the problem is even worse, because the underlying data for the case report is almost never made available to the journal or the reviewers; it is never even requested.

If a forensic expert bases his/her opinion on a case report or case series, but never actually reviews the underlying data, the expert's opinion is vulnerable to challenge. What if the author of the report has done a sloppy job, or believed their theory so strongly that they omitted all data that did not support the position they wished to promote, or wanted to get grant money for further research, or get more jobs as an expert witness? An expert who relies upon a case report or series of case reports, has based their opinion upon unverifiable data, unseen by reviewer or reader. This is hearsay.

4.3. Fraud

The possibility of research and publication fraud is well-recognised. Since 1994, when *bmj.com* began, the *BMJ* has retracted three studies and one letter, all fraudulent. Last June Malcolm Pearce, a British gynaecologist, was removed from the medical register for fraud: he had published two papers in the *British Journal of Obstetrics and Gynaecology* describing work that had never taken place.²¹ In August of last year *British Medical Journal* (*BMJ*) and *The Lancet* both stated they questioned the veracity of research they had published on the protective effects of diet on the heart, written by Ram B. Singh, head of the Haldberg Hospital and Research Institute in Moradabad, India.²² Articles in both journals cited the difficulties inherent in investigating scientific research fraud, and the even greater conundrums faced by journals forced to retract studies once they had been published.

4.4. Mistakes

Not all the misleading material published in medical journal is the product of fraudulent minds. Mistakes are made, and sometimes journal reviewers and editors fail to notice them. In October of last year a study showing that

the outcome of pregnancy in diabetic women in northeast England was worse than that of diabetic women in Norway, was retracted when the authors realized that they made a fundamental mistake in data collection.²³

4.5. Peer review failure

The uproar over VioxxTM at the NEJM is not the first problem that journal has had with the cardiovascular drug complications. In the year 2000 NEJM published a case series on the alleged cardiovascular toxicity of ephedrine-containing products.²⁴ The paper was based on a review of AERs about ephedrine-containing products received by the US Food and Drug Administration. Though not mentioned in the NEJM publication, the two authors of the paper had been paid “six to seven thousand dollars” each by the FDA (which wanted the drug outlawed) for writing the first report and, at the time of publication, one of the authors was negotiating with a plaintiff's attorney to act as an expert on one of the cases discussed in the paper.^A

The NEJM paper contained two tables, one listing cases where the authors had concluded that the cardiovascular complications described in the AERs were definitely, and the other probably, related to ephedra usage. Each table contained a column headed “preexisting conditions”, presumably to help readers decide for themselves whether the event reported might have been related to other causes besides ephedra.

In toxic tort litigation, when experts for the defense have shown that the injury in question could have been the result of established risk factors, and not the drug in question, the opposing expert must explain why he/she thinks that the drug, and not the recognized risk factors, was the cause of injury. The plaintiff's expert cannot just say that based on their experience, the drug caused the injury (*ipse dixit*). So it is important to know if other, more compelling, independent risk factors exist; could a pre-existing disease have caused the problem?

It seems quite uncontroversial that readers of the NEJM article would have expected conditions such as lethal cardiac malformations, cerebral aneurysm, pre-existing cardiac arrhythmias, smoking, diffuse coronary artery disease, occlusive carotid disease, drug abuse, or hyperthermia after pedaling a bicycle in a sauna while wearing a rubber suit, to be entered in the “pre-existing” column. All of these conditions were present, but never mentioned, presumably because the authors thought they were irrelevant. Any one of these conditions could have been the sole cause of death, but they were not mentioned in the “preexisting” columns. Worse the AER identifier numbers, which had been present in the original AERs provided by the FDA,

^A Copies of the original 140 adverse events report analyzed in the NEJM article, as well as a finders guide allowing readers to match patients described in the two tables with the original FDA identifiers, and a deposition given by one of the authors are posted on JCFM's website.

Score	Yes	No	Do not	Know
1. Are there previous conclusive reports of this reaction?	+1	0	0	
2. Did the adverse event occur after the suspected drug was given?	+2	-1	0	
3. Did the reaction improve when the drug was stopped?	+1	0	0	
4. Did the adverse reaction reappear when the drug was given again?	+2	-1	0	
5. Are there alternative causes that, on their own, could have caused the reaction?	-1	+2	0	
6. Did the reaction reappear when placebo was given?	-1	+1	0	
7. Was the blood level detected known to be toxic?	+1	0	0	
8. Was the reaction more severe when the dose was increased, or less severe when the dose was decreased?	+1	0	0	
9. Did the patient have a similar reaction to the same or similar drugs in any previous exposure?	+1	0	0	
10. Was the adverse event confirmed by any objective evidence?	+1	0	0	

Fig. 2. Sample Naranjo algorithm template.

were removed from the NEJM publication.²⁴ Their removal made it impossible for anyone reading the paper to check the case summaries for accuracy without devoting weeks to reading masses of documents. Thus, it seems likely that the reviewer or reviewers never saw the underlying data.

5. A possible solution

As the limitations of case reports have become more widely recognized, some journals have severely curtailed the number they are willing to publish. Other journals have adopted the use of pharmaco-vigilance ratings, the Naranjo algorithm being the most widely used template.²⁵ All of the different algorithms incorporate, some more than others, the original criteria put forward by Bradford Hill.²⁰

There are a number of pharmaco-vigilance schemes to choose from. None has been shown to be obviously superior to any of the others, but they all allow the reader to impose some structure on the evaluation of individual case reports. Having answered the standardized questions, a reviewer can then give a numerical score to each individual report, or consistently apply terms such as “unlikely”, “possible”, and “highly probable, etc.”. In and of itself, the grading process adds no further data, and it does nothing to validate or refute the nexus being suggested by the authors of the case report. It just makes the merits and faults stand out clearly for all to see.

When the standard Naranjo template (see Fig. 2) is applied to the first ephedrine case-report summarized in Section 2, the score is 0 (scores can be less than one). In the second case, the score was 3, considered a “possible”. In the Naranjo scoring system, case reports with scores of 9 are considered as strong evidence of a drug-related effect. When scores are 0 or less, any connection between exposure to the drug, and the observed effect, is considered doubtful.

If a journal is going to publish a case report about a drug reaction, it would seem reasonable to include the Naranjo score in the body of the report (the front matter of the journal could contain a brief explanation of how the system works – the *Journal American Family Practice* publishes its grading system for RCTs in each issue, and the *Annals of Pharmacotherapy* includes the Naranjo score with the case

report). Alternatively, journal editors might require the inclusion of a Naranjo score when the paper is first submitted. They might also ask for some documentation of the underlying data.

The proposed modifications in case reporting would require little effort and no expenditure, but they would be a first step towards curtailing some of the abuses of case reports now seen in the courts. A much more definitive solution would be to establish a forensics section of Cochrane Reviews, and preliminary steps in that direction are now being taken. The implementation of a Cochrane section might make the practice of forensic medicine more rational, but it will never entirely solve the problem. For one thing, it is not just young doctors that need training in the relative strengths of scientific evidence. The current generation of judges acting as gatekeepers are equally ill equipped for the task. Curriculum changes are needed in both law and medical schools.

Disclosure: I do not have a direct conflict of interest in relation with this letter, or the article to which it refers, but in the past I have received pay from numerous herbal supplement manufacturers, some of whom sell ephedrine.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jcfm.2006.02.022](https://doi.org/10.1016/j.jcfm.2006.02.022).

References

- DeBaakey L, DeBaakey S. The case report. I. Guidelines for preparation. *Int J Cardiol* 1983;4(3):357–64.
- Kelly W. The quality of published adverse event reports. *Ann Pharmacother* 2003;37(12):1774–8.
- Isner JM, Estes III NA, Thompson PD, et al. Acute cardiac events temporally related to cocaine abuse. *New Engl J Med* 1986;315(23):1438–43.
- Hollingsworth JG, Lasker EG. The case against differential diagnosis: Daubert, medical causation testimony, and the scientific method. *J Health Law* 2004;37(1):85–111.
- Libman RB, Menna BL, Gulati S. Case report: consequences of ephedra use in an athlete. *Lancet* 2005;366(Suppl. 1):S22.
- Chen-Scarabelli C, Hughes SE, Landon G, et al. A case of fatal ephedra intake associated with lipofuscin accumulation, caspase activation and cleavage of myofibrillary proteins. *Eur J Heart Fail* 2005;7(5):927–30.

7. Szakacs JE, Dimmette RM, Cowart Jr EC. Pathologic implication of the catechol amines, epinephrine and norepinephrine. *US Armed Forces Med J* 1959;**10**:908–25.
8. Szakacs JE, Cannon A. L-Norepinephrine myocarditis. *Am J Clin Pathol* 1958;**30**(5):425–34.
9. Kuhl U, Pauschinger M, Bock T, et al. Parvovirus B19 infection mimicking acute myocardial infarction. *Circulation* 2003;**108**(8): 945–50.
10. Hardman J, Limbird L, Goodman A. Pharmacokinetics. In: Wilkinson G, editor. *Goodman and Gilman's the pharmacological basis of therapeutics*. 10th ed. New York, Chicago, San Francisco, Lisbon, London, Madrid, Mexico City, Milan, New Delhi, San Juan Seoul, Singapore, Sydney, Toronto: McGraw-Hill Medical Publishing Division; 2001. p. 3–30.
11. Libman R, Menna B, Gulati S. Case report: consequences of ephedra use in an athlete. *Lancet* 2005;**366**(Suppl. 1):S22.
12. Morgenstern LB, Viscoli CM, Kernan WN, et al. Use of Ephedra-containing products and risk for hemorrhagic stroke. *Neurology* 2003;**60**(1):132–5.
13. Ariagno R, Karch SB, Middleberg R, Stephens BG, Valdes-Dapena M. Methamphetamine ingestion by a breast-feeding mother and her infant's death: People v Henderson. *JAMA* 1995;**274**(3):215.
14. Anon. Mother convicted for breast milk drug death of 3-month-old child. San Francisco Chronicle; 2003.
15. Stewart JL, Meeker JE. Fetal and infant deaths associated with maternal methamphetamine abuse. *J Anal Toxicol* 1997;**21**(6):515–7.
16. Anon. *Patterns of medication use in the United States, 2004*. Boston: Sloan Epidemiology Center, Boston University; 2004.
17. Association AH. American Heart Association's Heart Disease and Stroke Statistics – 2004 update. Dallas 2004.
18. Wadman M. Journal grows suspicious of Vioxx data. *Nature* 2005;**438**:899.
19. Curfman G, Morrissey S, Drazen J. Expression of concern: Bombardier et al., “Comparison of Upper Gastrointestinal Toxicity of Rofecoxib and Naproxen in Patients with Rheumatoid Arthritis”. *New Engl J Med* 2000;**343**:1520–8; *NEJM* 2005;**353**:26.
20. Bradford-Hill A. The environment and disease: association or causation? President's address. *Proc Roy Soc Med* 1965;**9**: 295–300.
21. Dryer O. Consultant struck off for fraudulent claims. *BMJ* 2005;**310**:1554–5.
22. White C. Suspected research fraud: difficulties of getting at the truth. *BMJ* 2005;**331**(7511):281–8.
23. Hawthorne G, Irgens L, Lie R, et al. Retraction of paper on maternal diabetes. *BMJ* 2003;**327**:7420–9.
24. Haller CA, Benowitz NL. Adverse cardiovascular and central nervous system events associated with dietary supplements containing ephedra alkaloids. *New Engl J Med* 2000;**343**(25):1833–8.
25. Naranjo CA, Busto U, Sellers EM, et al. A method for estimating the probability of adverse drug reactions. *Clin Pharmacol Ther* 1981;**30**(2):239–45.